UNIVERSITÉ DU LUXEMBOURG

Introduction à l'analyse factorielle dynamique

Jang Schiltz

Contexte expérimental

Recherche en situation naturelle

- Observer un individu sur de nombreux jours
- Observer un groupe d'individus une seule fois

Dimension temporelle et développementale fondamentale dans mesure des conduites !

Limitation des méthodes classiques

UNIVERSITÉ DU LUXEMBOURG

Analyse de séries temporelles d'un seul individu

$\Rightarrow$ nécessite beaucoup de mesures répétées !

Analyse multivariée classique

$\Rightarrow$ nécessite beaucoup d'individus!

Problème pour interpréter correctement de nombreuses études longitudinales effectués.

Analyse factorielle classique

Développée pour étudier les résultats de tests d'intelligence.

$$\boxed{y(t) = A\eta(t) + \varepsilon(t)}$$

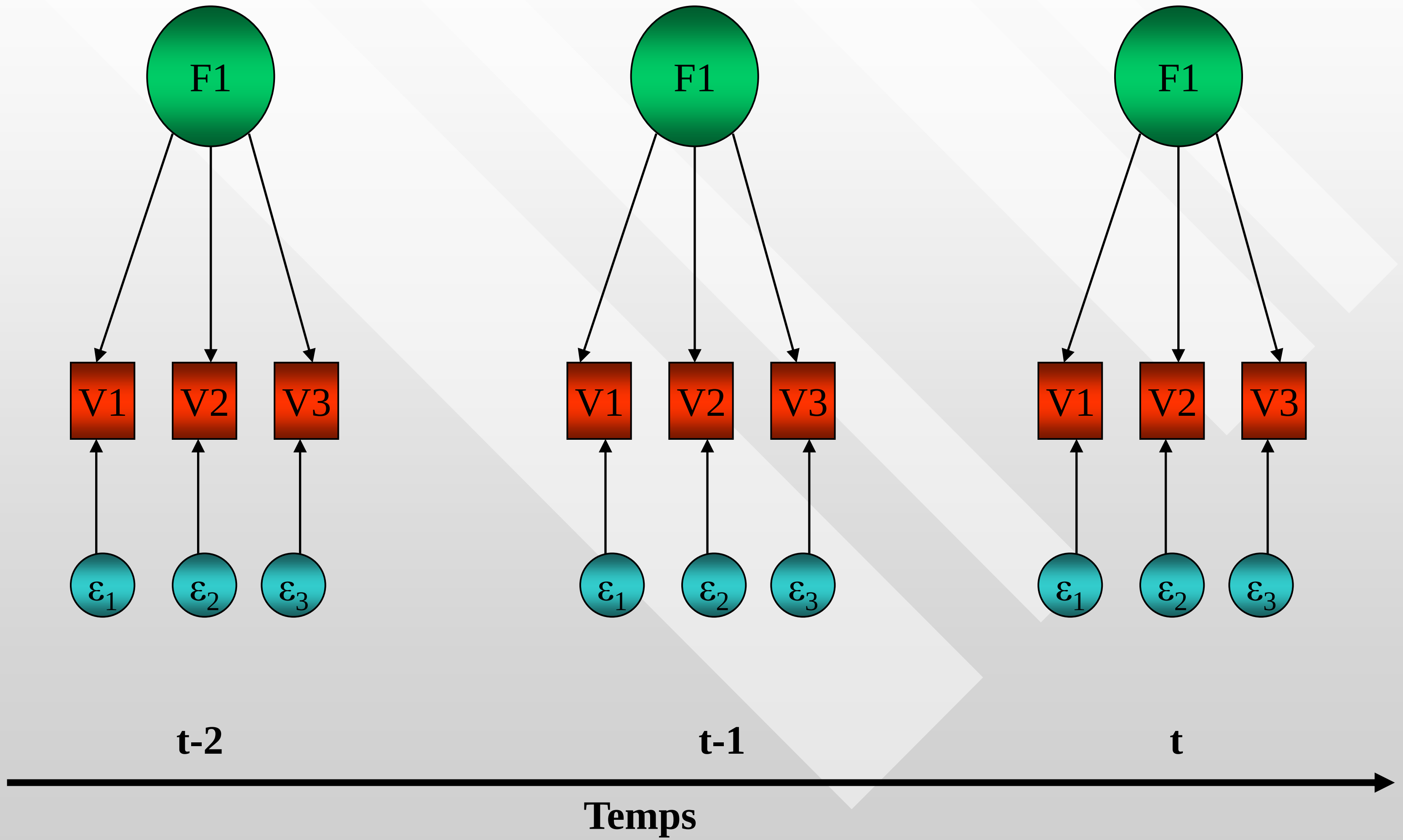
$y(t)$ est un vecteur $(n \times 1)$ représentant les n variables observées

$\eta(t)$ un vecteur $(k \times 1)$ représentant les k facteurs

A la matrice $(n \times k)$ comportant les saturations des n variables dans les k facteurs

$\varepsilon(t)$ un vecteur $(n \times 1)$ modélisant les erreurs de mesure

Analyse factorielle classique : schéma



AFD : 1er modèle

$$\boxed{y(t) = \Lambda(0)\cdot\eta(t) + \Lambda(1)\cdot\eta(t-1) + \cdots + \Lambda(s)\cdot\eta(t-s+1) + \varepsilon(t)}$$

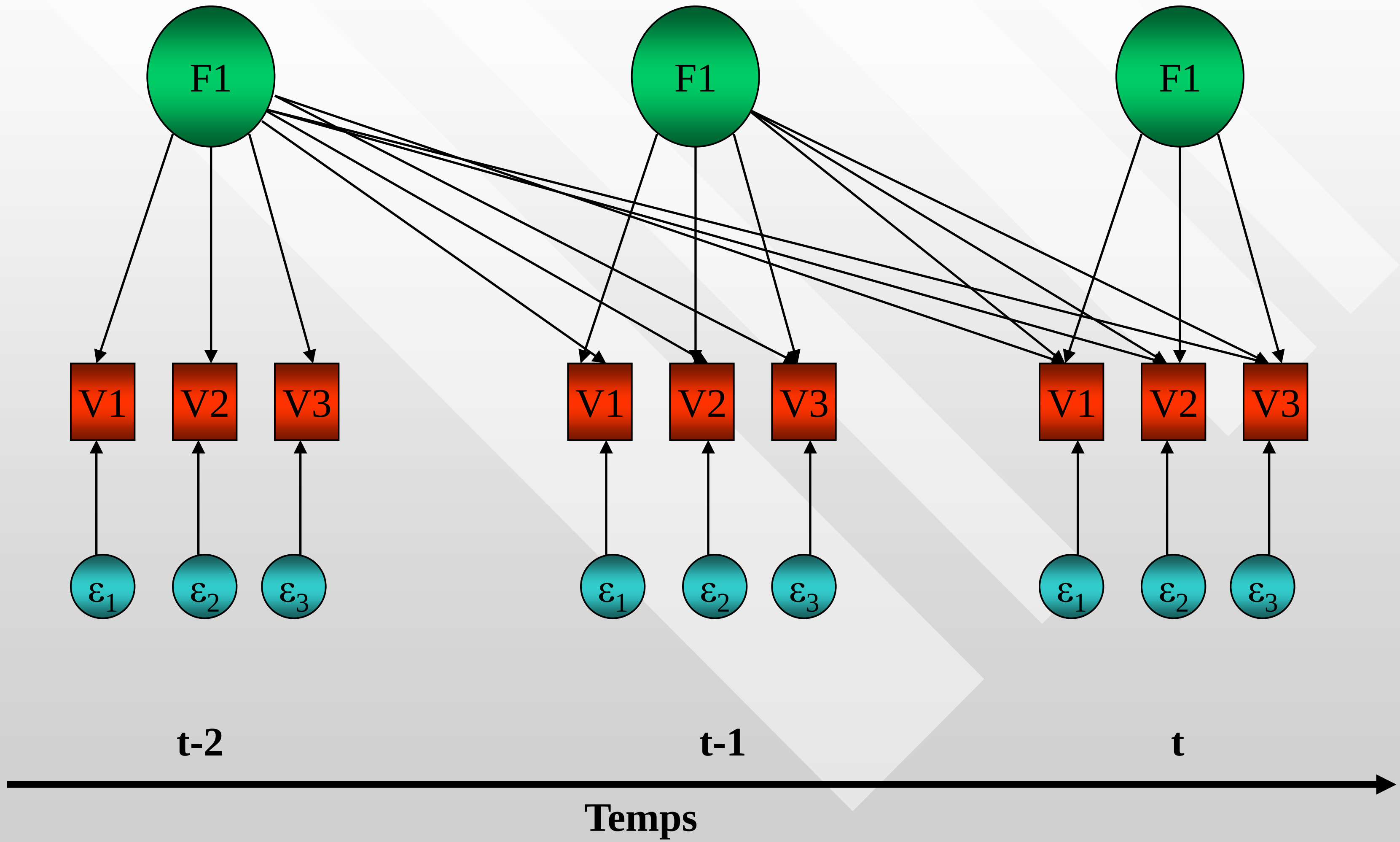
$y(t)$ est un vecteur $(n\times 1)$ représentant les n variables observées

$\eta(u)$ des vecteurs $(k\times 1)$ représentant les k facteurs qui expliquent les variables observés à l'instant u

$\Lambda(u)$ des matrices $(n\times k)$ comportant les saturations des n variables dans les k facteurs à l'instant u

$\varepsilon(t)$ un vecteur $(n\times 1)$ modélisant les erreurs de mesure

AFD 1er modèle : Schéma



AFD : 2e modèle

$$\boxed{y(t) = \Lambda\, f(t) + \varepsilon_1(t),}$$

$$\boxed{f(t) = B_1 f(t-1) + B_2 f(t-2) + \ldots + B_s f(t-s) + \varepsilon_2(t),}$$

$y(t)$ est un vecteur $(n \times 1)$ représentant les n variables observées

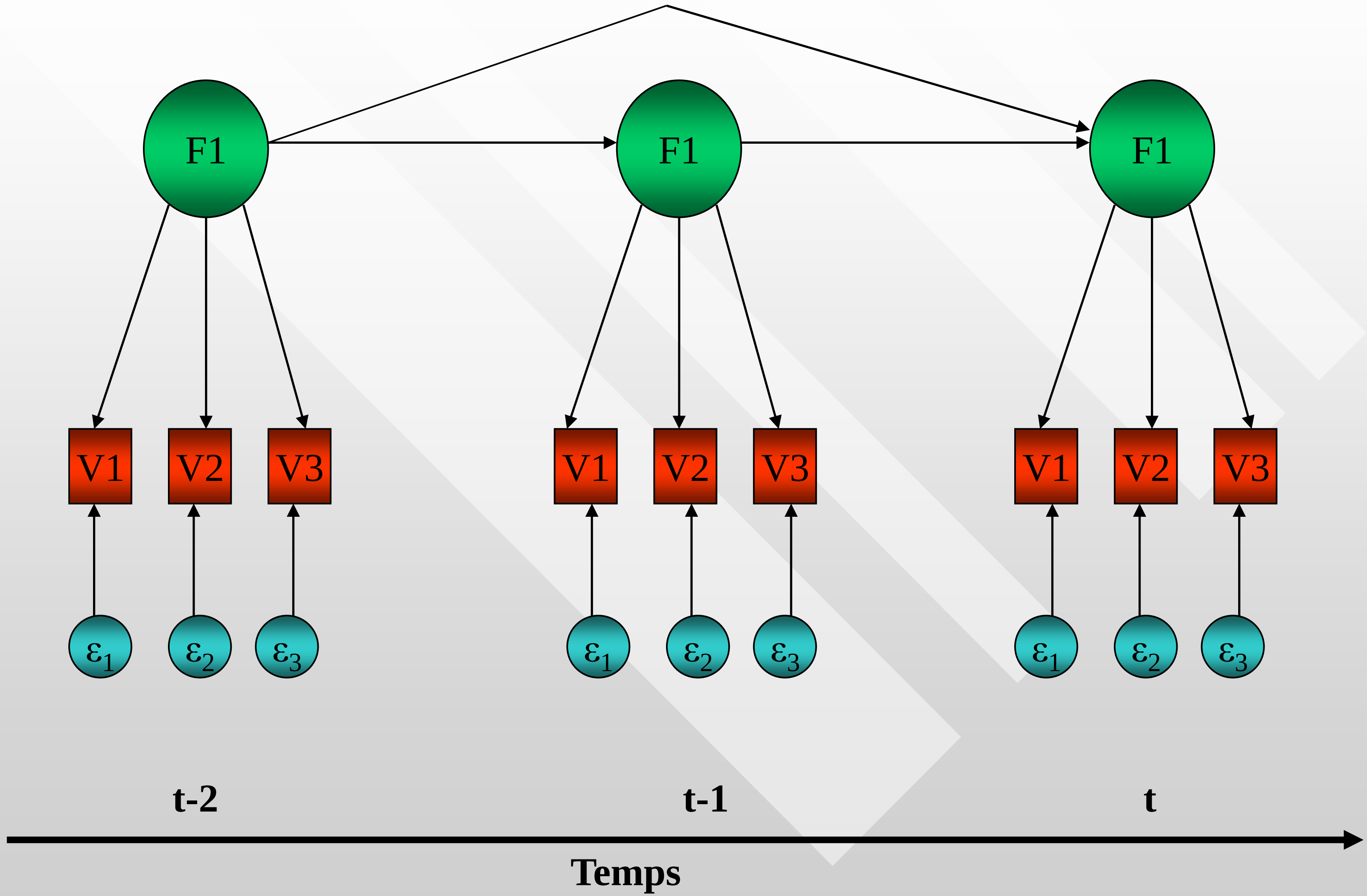
Λ est une matrice $(n \times k)$ comportant les saturations des n variables dans les k facteurs

$\varepsilon_1(t)$ et $\varepsilon_2(t)$ des vecteurs modélisant les erreurs de mesure

$f(u)$ des vecteurs $(k \times 1)$ représentant les k facteurs qui expliquent les variables observés à l'instant u

B_l sont des matrices $(n \times k)$ qui permettent de modéliser les corrélations autorégressifs respectivement régressifs en croix

AFD : 2e modèle : Schéma



Avantages de l'AFD

Analyse des séries temporelles d'une seule personne :

plusieurs centaines de mesures

Analyse factorielle dynamique :

20 à 50 mesures d'une dizaine de personnes

Beaucoup plus facile à réaliser expérimentalement

Mais il faut avoir le droit d'agréger les données !!!

Les matrices de covariances des facteurs doivent être « proches »

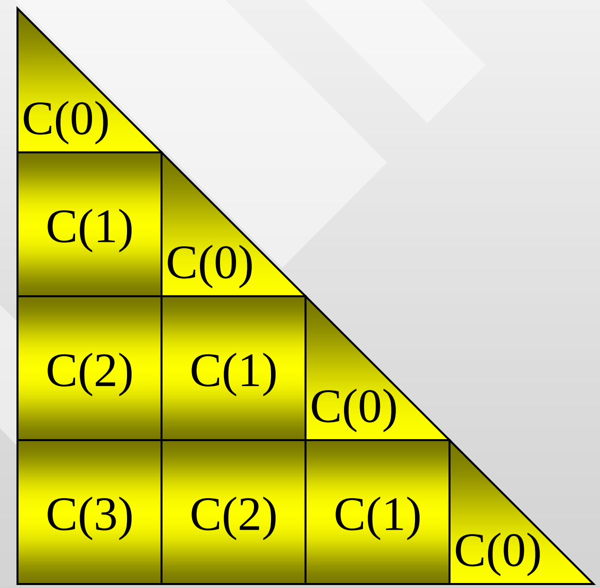
Construction de la matrice de Toeplitz

Notons $y_i(t)$ les variables observées de la i-ème personne à l'instant t

$$C_i(u) = \mathrm{Cov}\big(y_i(t), y_i(t+u)\big)$$

$C_i(0)$ est la matrice de covariance habituelle pour la *i*-ème personne qu'on factorise dans l'analyse factorielle classique.

$$T_i(s) = C_i(j-k), \quad j,k = 0,1,\ldots,s$$
$$C_i(-u) = C_i^t(u).$$



Matrice de Toeplitz pour s = 3

Test d'agrégation

Test du khi-deux à $[1/2k(k+1)+sk^2](N-1)$ degrés de libertés sur l'ensemble des matrices de Toeplitz.

Que faire si on n'a pas le droit d'agréger les données des N personnes?

1) Éliminer celui qui a la plus grande valeur de khi-deux et refaire le test avec les autres.

2) Calculer les valeurs du khi-deux pour tous les sous-ensembles à N-1 personnes.

Retenir le sous-ensemble avec la plus petite valeur s'il passe le test.

Si aucun sous-ensemble à N-1 personnes passe le test, prendre tous les sous-ensembles à N-2 personnes, etc.

Conclusion

- L'analyse factorielle dynamique est par conséquent une technique qui permet de détecter des structures communes dans des séries temporelles, ainsi que les relations entre les séries observées et les variables explicatives.

- C’est une méthode utilisée tout aussi bien en économétrie et économie qu’en psychologie et sciences sociales.